A practical procedure to find matching priors for frequentist inference

By JUAN ZHANG and JOHN E. KOLASSA

Department of Statistics, Rutgers University,

Piscataway, New Jersey, 08854, U.S.A.

janezh@stat.rutgers.edu kolassa@stat.rutgers.edu

Summary

We present a practical way to find the matching priors proposed by Welch & Peers (1963) and Peers (1965). We investigate the use of saddlepoint approximations combined with matching priors and obtain p-values of the test of an interest parameter in the presence of nuisance parameter. The advantage of our procedure is the flexibility of choosing different initial conditions so that one can adjust the performance of the test. Two examples have been studied, with coverage verified via Monte Carlo simulation. One relates to the ratio of two exponential means, and the other relates the logistic regression model. Particularly, we are interested in small sample settings.

Some key words: Bayes; Conditional inference; Matching prior; Modified signed root likelihood ratio statistic; Partial differential equation; Saddlepoint approximation.

1. Introduction

We consider inference on a single scalar parameter in the presence of nuisance parameters. Under the frequentist settings, conditional inference can be complicated. Bayesian method can simplify frequentist elimination of nuisance parameters. The frequentist and the Bayesian approaches can be connected by matching priors. Matching priors were first proposed by Welch & Peers (1963) and Peers (1965). Determining a matching prior is equivalent to finding a solution of a first order partial differential equation. Only in simple circumstances, such as when parameters are orthogonal, the partial differential equation can be solved analytically. Levine & Casella (2003) note that "Unfortunately, except for these cases, the solution of the resulting partial differential equations becomes quite a hurdle; our only hope is to find numerical solutions to these partial differential equation."

We will see a practical way to solve for the matching priors, without the involvement of the back transformation described by Levine & Casella (2003). This procedure is easy to understand, can be implemented in R, R Development Core Team (2007) and is suitable to all kinds of initial conditions.

Our implementation of matching priors for the approximations proposed by DiCiccio & Martin (1993) is less complicated than other frequentist methods. DiCiccio and Martin's approximations are saddlepoint approximations that make use of Bayesian–frequentist parallels. Our proposed implementation requires less computational effort compared to the iterative Metropolis-Hasting algorithm described by Levine & Casella (2003).

We end the introduction with a brief outline of this paper. In §2, We review the concepts of matching priors and discuss the circumstance when orthogonal parameters are presence. Existing analytical and numerical solutions are reviewed.

In §3, we present the procedure for solving matching priors, both analytically and numerically. Specification of initial condition is discussed. We also provide information of R software implementation of the solving procedure. In §4, the approximations of DiCiccio & Martin (1993) are reviewed. The application of using matching priors conjuncted with DiCiccio and Martin's approximations is illustrated through examples in section §5. Different initial conditions are specified for obtaining various matching priors. Finally, §6 contains the conclusion.

2. Matching priors

We consider parametric models with random variables $X_1, \ldots, X_n$ having joint density function that depends on the unknown parameter vector ω. Suppose ω is of length d and $\omega = (\omega^1, \omega^2, \ldots, \omega^d) = (\psi, \lambda)$ with $\psi = \omega^1$, the parameter of interest, and the nuisance parameter $\lambda = (\omega^2, \ldots, \omega^d)$.

Matching priors were proposed by Welch & Peers (1963) and Peers (1965). In the following, denote the matching prior by $\pi(\cdot)$. Let $\mathrm{pr}_\pi(\cdot|X)$ be the posterior probability measure for ψ under prior $\pi(\cdot)$. The upper $(1-\alpha)$ posterior quantile constructed on the basis of a prior density function $\pi(\psi)$ has the property that it is also the frequentist limit, such that

$$\mathrm{pr}_\pi\{\psi \le \psi^{(1-\alpha)}(\pi, X)|X\} = \mathrm{pr}_\psi\{\psi \le \psi^{(1-\alpha)}(\pi, X)\} = 1 - \alpha + O(n^{-1}).$$

When there are no nuisance parameters, Welch & Peers (1963) showed that the appropriate choice of $\pi(\omega)$ is $\pi(\omega) \propto \{i(\omega)\}^{1/2}$, where $i(\omega) = \mathrm{E}\{-\mathrm{d}^2\, l(\omega)/\,\mathrm{d}\,\omega^2\}$, and $l(\cdot)$ is the log-likelihood function. In this case, matching priors can be easily obtained.

In the presence of nuisance parameters, Peers (1965) showed that $\pi(\omega)$ must

be chosen to satisfy the partial differential equation

$$\sum_{j=1}^{d} i^{1j}(i^{11})^{-1/2}\frac{\partial}{\partial\omega^j}(\log\pi) + \sum_{j=1}^{d}\frac{\partial}{\partial\omega^j}\{i^{1j}(i^{11})^{-1/2}\} = 0, \tag{1}$$

where $i_{jk}(\omega) = \mathrm{E}\{-\partial^2 l(\omega)/\partial\omega^j\partial\omega^k\}$ and (i^{jk}) is the $d\times d$ inverse matrix of (i_{jk}).

If the parameter of interest and the nuisance parameter vector are orthogonal, solving the partial differential equation (1) is relatively easy. We follow the definition of parameter orthogonality by Cox & Reid (1987). Orthogonality is defined with respect to the expected Fisher information matrix. The most direct statistical interpretation of parameter orthogonality is that the relevant components of the original statistic are uncorrelated. In general, it is possible to obtain orthogonality of a scalar parameter of interest to a set of nuisance parameters.

When the parameter of interest ψ is orthogonal to a set of nuisance parameters, equation (1) reduces to

$$(i_{\psi\psi})^{-1/2}\frac{\partial}{\partial\psi}(\log\pi) + \frac{\partial}{\partial\psi}(i_{\psi\psi})^{-1/2} = 0. \tag{2}$$

Tibshirani (1989) showed that solutions were of the form $\pi(\psi,\lambda) \propto \{i_{\psi\psi}(\psi,\lambda)\}^{1/2}$ $g(\lambda)$, where $g(\lambda)$ is arbitrary, and the suggestive notation $i_{\psi\psi}(\psi,\lambda)$ is used in place of $i_{11}(\psi,\lambda)$.

However, choosing a parametrization to achieve parameter orthogonality is not always easy, and it can be hard in some cases. It is equivalently hard to obtain orthogonalization and to solve the partial differential equation (1) directly, since the orthogonalization procedure also requires solutions to partial differential equations of form similar to (1). Staicu and Reid (2007), studied the use of matching priors with the approximation of DiCiccio & Martin (1993) under orthogonal parametrization, and showed that the Peers-Tibshirani class of matching priors was essentially unique. One can modify the arguments in this paper to solve the

partial differential equation that defines the orthogonality transformation, and attempt, using orthogonality, to narrow down the class of matching priors.

Levine & Casella (2003) proposed a general procedure to solve the partial differential equation (1) numerically, in models with a single nuisance parameter. Firstly, they transform the parameters into another parameter space, solving the equation, and then transform back to the original parameter space. The numerical application of this procedure is not necessarily easy, and usually the transformation between the two parameter spaces is nontrivial. Levine & Casella (2003) implemented their procedure using Mathematica. They did not give instructions on initial condition specification, which is a necessary component to give specific solution in solving the partial differential equation. Sweeting (2005) introduced data-dependent priors that locally approximate the matching priors, and his procedure can deal with vector nuisance parameters.

3. Solving for the matching priors

In this section, we introduce a procedure to solve the partial differential equation (1) in general parametrization. For simplicity, we consider the dimension of the parameter space as 2. First, we give analytical form of the solutions, and then practical notes will be presented later in this section.

In the case that $d = 2$, equation (1) is reduced to

$$a(\psi, \lambda) z_\psi + b(\psi, \lambda) z_\lambda = d(\psi, \lambda), \tag{3}$$

where

$$z(\psi, \lambda) = \log\{\pi(\psi, \lambda)\},$$

$$a(\psi, \lambda) = \{i^{11}(\psi, \lambda)\}^{1/2},$$

$$b(\psi, \lambda) = i^{12}(\psi, \lambda)\{i^{11}(\psi, \lambda)\}^{-1/2},$$

and

$$d(\psi,\lambda) = -\left[\frac{\partial}{\partial\psi}\{i^{11}(\psi,\lambda)\}^{1/2} + \frac{\partial}{\partial\lambda}\{i^{12}(\psi,\lambda)\}\{i^{11}(\psi,\lambda)\}^{-1/2}\right].$$

The coefficient $a(\psi,\lambda)$ is a diagonal element of the inverse matrix of (i_{jk}), so $a(\psi,\lambda)$ can not be zero. Dividing both sides of (3) by $a(\psi,\lambda)$, we have

$$z_\psi + \frac{b(\psi,\lambda)}{a(\psi,\lambda)} z_\lambda = \frac{d(\psi,\lambda)}{a(\psi,\lambda)}.$$

This forces the coefficient of z_ψ to be 1, which simplifies the procedure of finding a solution.

To solve the equation (1), it suffices to solve the following ordinary differential equations system

$$\frac{\mathrm{d}\,\psi}{\mathrm{d}\,s} = 1, \frac{\mathrm{d}\,\lambda}{\mathrm{d}\,s} = \frac{b(\psi,\lambda)}{a(\psi,\lambda)}, \frac{\mathrm{d}\,z}{\mathrm{d}\,s} = \frac{d(\psi,\lambda)}{a(\psi,\lambda)}. \tag{4}$$

To be more specific with the solution, let us consider the initial conditions prescribed along an initial curve I. Suppose that I is given parametrically, in terms of a parameter ξ, as

$$\psi = \Psi(\xi), \quad \lambda = \Lambda(\xi).$$

Then evaluating $z(\psi,\lambda)$ at a point on I is equivalent to expressing z as a function of ξ,

$$z = Z(\xi) = z\{\Psi(\xi), \Lambda(\xi)\}. \tag{5}$$

Here, it is obvious to see that I can not be tangent to the direction $\left[1, \frac{b\{\Psi(\xi),\Lambda(\xi)\}}{a\{\Psi(\xi),\Lambda(\xi)\}}\right]$. We then obtain

$$\psi = \psi(s,\xi), \quad \lambda = \lambda(s,\xi)$$

by simultaneously integrating the two equations defined by

$$\frac{\mathrm{d}\,\psi}{\mathrm{d}\,s} = 1, \quad \psi(s_0,\xi) = \Psi(\xi), \tag{6}$$

$$\frac{\mathrm{d}\,\lambda}{\mathrm{d}\,s} = \frac{b(\psi,\lambda)}{a(\psi,\lambda)}, \quad \lambda(s_0,\xi) = \Lambda(\xi). \tag{7}$$

From the third equation in (4), the initial condition is given by (5). Then we have,

$$\frac{\mathrm{d}\, z}{\mathrm{d}\, s} = \frac{d(\psi, \lambda)}{a(\psi, \lambda)}, \quad z(s_0, \xi) = Z(\xi), \tag{8}$$

Equation (8) can be integrated by quadrature, once equations (6) and (7) have been solved,

$$z(s, \xi) = Z(\xi) + \int_{s_0}^{s} \frac{d\{\psi(s', \xi), \lambda(s', \xi)\}}{a\{\psi(s', \xi), \lambda(s', \xi)\}} \,\mathrm{d}\, s'. \tag{9}$$

These generate a surface in three dimensions, $Z(\psi, \lambda)$, that satisfies both the equation (3) and the initial condition. When there are no close form solutions for equations (6),(7) and (8), numerical solutions can be achieved. Rhee et al. (1986) presents more mathematical details.

In obtaining the solution formula (9) of $z(s, \xi)$, we avoid doing back transformation as described by Levine & Casella (2003). Noticing that if we want to specify the value of a matching prior at a certain point, say (ψ^*, λ^*), we can directly specify s as ψ^* and ξ as λ^* in formula (9), and then the matching prior evaluated at (ψ^*, λ^*) can be achieved.

Without loss of generality, set the initial condition

$$\{\Psi(\xi), \Lambda(\xi), Z(\xi)\} = (0, \xi, -1).$$

With $\Psi(\xi) = 0$, we have $\psi = s$. The equations (7) and (8) can be simplified as

$$\begin{aligned} \frac{\mathrm{d}\, \lambda}{\mathrm{d}\, s} &= \frac{b(s, \lambda)}{a(s, \lambda)}, \quad \lambda(s_0, \xi) = \Lambda(\xi), \\ \frac{\mathrm{d}\, z}{\mathrm{d}\, s} &= \frac{d(s, \lambda)}{a(s, \lambda)}, \quad z(s_0, \xi) = Z(\xi). \end{aligned} \tag{10}$$

We used R package `odesolve` by Setzer (2007) to solve equation (10) and get a numerical expression of $\lambda(\cdot)$ in s. The command `lsoda()` in `odesolve` package is designed to solve initial value problems for stiff or non-stiff systems of first order ordinary differential equations. It provides an interface to the Fortran ordinary

differential equation solver of the same name, written by Hindmarsh (1983) and Petzold (1983). For (9), we did numerical integration using Simpson's Rule and employed the R function `sintegral()` in the `Bolstad` package by Curran (2005). Suppose z will be evaluated at (ψ^*, λ^*). Noticing that $\Lambda(\xi) = \xi$, choose the start value as λ^* in solving (10), and then choose the upper integration limit as ψ^* in (9). The procedure is easy to perform if one has an ordinary differential equation solver, even if not using the solver provided by R package `odesolve`.

Based on the ordinary differential equation (6), $\psi = s+\Psi(\xi)$, i.e. $s = \psi-\Psi(\xi)$. So s_0 must be chosen considering the range of ψ. If we choose $\Psi(\xi) = 0$, then $\psi = s$. For the first example in §5, the parameter ψ is the ratio of two exponential means, and hence $\psi > 0$. Therefore, s_0 should be chosen as any positive value.

In the above we choose the initial values as $\{\Psi(\xi), \Lambda(\xi), Z(\xi)\} = (0, \xi, -1)$. Now we will show that the numerical solving procedure is suitable to any initial values.

• Suppose the initial condition for the ordinary differential equation (7) is $\lambda(s_0, \xi) = \Lambda(\xi)$, for $\Lambda(\xi)$ an arbitrary known function rather than $\Lambda(\xi) = \xi$ as above. The solution formula of z is the same as stated in (9). When solving (7), the initial value should be chosen as $\Lambda(\lambda^*)$, no longer λ^*, if z is evaluated at (ψ^*, λ^*).

• If the initial condition of (6) is $\psi(s_0, \xi) = \Psi(\xi)$, then the solution from the equation (6) is $\psi = s + \Psi(\xi)$. Therefore, the equation (7) becomes,

$$\frac{\mathrm{d}\,\lambda}{\mathrm{d}\,s} = \frac{b\{s + \Psi(\xi), \lambda\}}{a\{s + \Psi(\xi), \lambda\}}.$$

Let $\tilde{s} = s + \Psi(\xi)$. By simple change of variables, (7) becomes

$$\frac{\mathrm{d}\,\lambda}{\mathrm{d}\,\tilde{s}} = \frac{b(\tilde{s}, \lambda)}{a(\tilde{s}, \lambda)}.$$

Equation (8) is

$$\frac{\mathrm{d}\,z}{\mathrm{d}\,\tilde{s}} = \frac{d[\psi\{\tilde{s} - \Psi(\xi), \xi\}, \lambda\{\tilde{s} - \Psi(\xi), \xi\}]}{a[\psi\{\tilde{s} - \Psi(\xi), \xi\}, \lambda\{\tilde{s} - \Psi(\xi), \xi\}]}$$

with $z\{\tilde{s}_0 - \Psi(\xi), \xi\} = Z(\xi)$, noticing that $\tilde{s}_0 = s_0 + \Psi(\xi)$. Then the solution of z is simply given by the following formula,

$$z(\tilde{s}, \xi) = Z(\xi) + \int_{s_0 - \Psi(\xi)}^{\tilde{s} - \Psi(\xi)} \frac{d\{\psi(s', \xi), \lambda(s', \xi)\}}{a\{\psi(s', \xi), \lambda(s', \xi)\}} \, \mathrm{d}\, s'. \tag{11}$$

That is to say, the value of the prior on a certain point with the initial condition $\psi(s_0, \xi) = \Psi(\xi)$, is obtained by translating the interval of integration when $\Psi(\xi) = 0$ by $\Psi(\xi)$.

• Suppose the initial condition for (8) is $z(s_0, \xi) = Z(\xi)$ and $Z(\cdot)$ is a known function. This case is even simpler to deal with. One only needs to plug the value of $Z(\xi)$ into (9).

Therefore, the suggested numerical solving procedure is suitable to any initial values.

In the above, both the parameter of interest and the nuisance parameter are scalars. With dimension 2, it is relatively easy to understand the first order partial differential equation solving procedure from the geometric point of view, since one can draw the initial conditions and the solution surface in a 3-dimensional space. In Zhang (2008), the solving procedure was extended to multiple nuisance parameters, while keeping the parameter of interest as a scalar. The procedure of the higher dimension is similar as the one of 2-dimensional model parameters. However, when $d > 2$, it can be computational intensive to implement the procedure. Also, if there are no explicit expressions for the coefficients in the original first order partial differential equation, numerical implementation may be more difficult.

4. DiCiccio and Martin's approximations

Likelihood ratio test is widely used in statistical inference. The signed root of the likelihood ratio statistic is $R = \mathrm{sgn}(\hat{\psi} - \psi_0)[2\{l(\hat{\omega}) - l(\psi_0, \hat{\lambda}_0)\}]^{1/2}$, where

$l(\omega)$ is the log-likelihood function for the unknown parameter vector ω and $\hat{\lambda}_0$ is shorthand for $\hat{\omega}_{\psi_0}$, the constrained maximum likelihood estimator of ω. The standard normal approximation to the distribution of R typically has error of order $O(n^{-1/2})$, and R can be used to construct approximate confidence limits for ψ having coverage error of order $O(n^{-1/2})$.

Using matching priors, DiCiccio & Martin (1993) proposed tail probability approximations of order $O(n^{-1})$. The approximations are saddlepoint approximations that involve Bayesian method. The approximations of DiCiccio & Martin (1993) can be expressed in the Barndorff-Nielsen (1980) format

$$\Phi\{R + R^{-1}\log(T/R)\}, \tag{12}$$

and the Lugannani & Rice (1980) format

$$\Phi(R) + \phi(R)(R^{-1} - T^{-1}), \tag{13}$$

where Φ is the standard normal distribution function, and T is defined as

$$T = l_{\psi}(\psi_0, \hat{\lambda}_0)\frac{|-l_{\lambda\lambda}(\psi_0, \hat{\lambda}_0)|^{1/2}\pi(\hat{\omega})}{|-l_{\omega\omega}(\hat{\omega})|^{1/2}\pi(\psi_0, \hat{\lambda}_0)}. \tag{14}$$

Here $l_{\psi}(\omega) = \partial l(\omega)/\partial\psi$, $l_{\omega\omega}$ is the matrix of second-order partial derivatives of $l(\omega)$ taken with respect to ω; $l_{\lambda\lambda}(\omega)$ is the submatrix of $l_{\omega\omega}(\omega)$ corresponding to λ; and $\pi(\omega)$ is a matching prior density for $\omega = (\psi, \lambda)$ which satisfies equation (1). Then the resulting approximation is $\mathrm{pr}(\psi \geq \psi_0|X) \doteq \Phi\{R + R^{-1}\log(T/R)\}$, or, $\mathrm{pr}(\psi \geq \psi_0|X) \doteq \Phi(R) + \phi(R)(R^{-1} - T^{-1})$. Both of them have relative error of order $O(n^{-1})$. Approximate confidence limits for ψ can be constructed using either of (12) or (13). These confidence limits have coverage errors of order $O(n^{-1})$. To relative error of the order $O_p(n^{-1})$, the variable T is parameterization invariant under transformations $\omega \mapsto \{\psi, \tau(\omega)\}$.

The approximations of DiCiccio & Martin (1993) show their advantages in less computational effort compared to the Metropolis-Hastings procedure used

by Levine & Casella (2003). To calculate T in (14), the matching prior requires to be evaluated at two points, $(\psi_0, \hat{\lambda}_{\psi_0})$ and $\hat{\omega}$. The initial curve can be chosen passing through $(\psi_0, \hat{\lambda}_{\psi_0})$; that is to say, only the solution on one point $\hat{\omega}$ needs to be determined.

5. Examples

5.1. *Ratio of two exponential means*

Let X and Y be exponential random variables with means μ and ν respectively; the ratio of the means, ν/μ, is the parameter of interest. The parameter transformation $\left(\mu \to \lambda\psi^{-\frac{1}{2}},\ \nu \to \lambda\psi^{\frac{1}{2}}\right)$ makes the two new parameters ψ and λ orthogonal. Then X and Y have expectations $\lambda\psi^{-\frac{1}{2}}$ and $\lambda\psi^{\frac{1}{2}}$, respectively.

Suppose we have n independent replications of (X, Y). Denote $\omega = (\psi, \lambda)$. We can obtain the log-likelihood function as $l(\omega) = -n\left\{(\psi\bar{x} + \bar{y})/(\lambda\sqrt{\psi}) + 2\log\lambda\right\}$.

Both approximations of the Barndorff-Nielson format (12) and the Lugannani and Rice format (13) are considered. Based on these approximations, p-values can be calculated. Approximations based on different prior density functions mentioned previously may be used to generate an approximate one-sided p-value by approximating $\mathrm{pr}(R \geq r)$, for r the observed value of R. Approximate two-sided p-values may be calculated by approximating $2\min\{\mathrm{pr}(R \geq r), \mathrm{pr}(R < r)\}$. One and two-sided hypotheses tests of size α may be constructed by rejecting the null hypothesis when the p-value is less than α. Table 1 reports type I error probabilities of the 1,000,000 rounds of simulation with $n = 10$.

In this example, the parameters ψ and λ are orthogonal. Using the simplified partial differential equation (2), $\pi(\psi, \lambda) = 1/\psi$ is an explicit solution. Also $\pi(\psi, \lambda) = 1/(\psi\lambda)$ is another explicit solution. Numerical solutions were also calculated. One of the initial condition is $\{\Psi(\xi), \Lambda(\xi), Z(\xi)\} = (0, \xi, -1)$. The

resulting matching prior corresponds to the the analytic solution $1/\psi$. Another numerically solved matching prior is based on the initial condition $(0, \xi, -\log \xi)$, which corresponds to the the analytic solution $1/(\psi\lambda)$. From Table 1, one can see that the numerical and analytic solutions give almost the same simulation results, which confirmed the validity of our numerical solution process.

Approximations (12) and (13) have a removable singularity at $R = 0$. Consequently, these and similar formulae require care when evaluating near $R = 0$. In these cases, for all but the most extreme conditioning events, the resulting conditional p-value is large enough as to not imply rejection of the null hypothesis, and so these simulated data sets are treated as not implying rejection of the null hypothesis.

5.2. *Logistic regression*

We consider a logistic regression model with a binary response Y and only one explanatory variable X. Let ω_1 denote the unknown intercept and ω_2 denote the unknown effect of the explanatory variable. Suppose ω_2 is the parameter of interest and ω_1 is the nuisance parameter. We will solve matching priors and apply DiCiccio and Martin's approximations to do inference about ω_2. Levine & Casella (2003) considered a similar example.

Let Y_i be the response variable taking binary values with success probability as p_i, and X_i be the explanatory variable following uniform distribution $U(0, 1)$. Suppose there are n independent replications of (X_i, Y_i). Fit the model $\log\{p_i/(1-p_i)\} = v_i'\omega = \omega_1 + \omega_2 x_i$, where $v_i = (1, x_i)'$ and ω_2 is the parameter of interest. Inverting the equation, we have $p_i = (1 + e^{-v_i'\omega})^{-1}$. We can obtain the log-likelihood function as $l(\omega; x) = \sum_{i=1}^{n} y_i \log\{p_i/(1-p_i)\} + \sum_{i=1}^{n} \log(1-p_i)$. The first derivative of the log=likelihood function is $V'(y-p)$, where V is the design matrix with v_i' in row i. The second derivative of log-likelihood function

is $-V'WV$, where W is a diagonal matrix with diagonal elements $p_i(1-p_i), i = 1, \cdots, n$.

Using sample size $n = 30$, we generate data satisfying the logistic regression model with $\omega_1 = -1$, $\omega_2 = 0.5$, and the explanatory variable X following uniform distribution $U(0, 1)$. In this case, generally the parameters ω_1 and ω_2 are not orthogonal. We use the numerical procedure described in §3 and study performances of different initial conditions. Table 2 contains type I error probabilities for both one-sided and two-sided tests for approximations of both Barndorff-Nielson format and Lugannani and Rice format, based on 10,000 rounds of simulation.

As we mentioned previously, approximations (12) and (13) have a removable singularity at $R = 0$. We deal with this singularity the same way as in §5.1.

In the following, we give some instructions on how to change the initial condition and how to choose favorable initial conditions. Initial condition $(0, \xi, -1)$ gives type I error probabilities larger than the nominal level 0.05; that is to say, it has the tendency to underestimate tail probabilities and reject the null hypothesis. We want to choose initial conditions to obtain a test whose type I error rate is closer to the nominal level. We adjust the initial condition when solving the partial differential equation (1), and use the Barndorff-Nielson format of the approximation. The quantity T in (14) is the only part in the approximation that relates to matching priors. For a one-sided test, when the probability is small and close to 0, R and T are negative. Making $\Phi\{R + R^{-1}\log(T/R)\}$ larger is equivalent to making T bigger. Also one may notice that $Z(\xi)$ is used only in equation (9). Suppose the initial condition is $\{\Psi(\xi), \Lambda(\xi), Z(\xi)\}$. Keep the first two components of the initial condition, $\Psi(\xi)$ and $\Lambda(\xi)$, unchanged, and only modify the third term, $Z(\xi)$. By doing so, the integral part in equation

(9) is kept unchanged and z varies only with $Z(\xi)$. By changing $Z(\xi)$, we want to adjust T to be bigger. Because T is negative when reject a hypothesis, and matching priors appear in T as a ratio, one can construct a $Z(\cdot)$ such that the ratio, $\exp\{Z(\hat{\psi}, \hat{\lambda})\}/\exp\{Z(\psi_0, \hat{\lambda}_0)\}$, will be smaller than 1; recall that 1 is the value of the ratio when $Z(\xi) = -1$. Based on the above arguments, $Z(\cdot)$ function is constructed as $Z(\xi) = -\log\{(\xi+1)^q + 1\}$, where q is a tuning parameter and leads $Z(\cdot)$ to an even function. As an even function, $Z(\xi)$ achieves its maximum value at -1, where -1 is the true value for the nuisance parameter when data were simulated. We have constructed priors using knowledge of the true value of the nuisance parameter. Of course, in practice this knowledge is unavailable. One might instead use an estimator of the nuisance parameter in place of the true value.

When $Z(\xi)$ increases quickly, such as $q = 2$ in table 2, the type I error probability deviates far away from the nominal level in the other direction. If a more slowly increasing functions is used, the performance of type I error may be better.

Unfortunately, with some choices of initial conditions, such as the last three listed in table 2, the Lugannani and Rice format approximation may fall outside the range of 0 and 1 in some cases. For example, the initial condition of $[0, \xi, -\log\{(\xi+1)^2 + 1\}]$ yielded 5 such probabilities out of 10,000 data sets. We convert those values to 0 or 1 by $\min\{\max(p, 0), 1\}$, where p is the p-value that is outside 0 and 1.

For the parameter of interest ω_2, we calculate credible intervals using DiCiccio and Martin's approximation in Barndorff-Nielson format. With initial condition $(0, \xi, -1)$, out of 1,000 generated data sets, there are 938 credible intervals covered the true value 0.5. With initial condition $[0, \xi, -\log\{(\xi+1)^{2/5} + 1\}]$, for the

parameter of interest ω_2, there are 954 credible intervals covered the true value 0.5.

We apply the above procedure to a real data set from Hosmer & Lemeshow (2000, Table 1.1). The response variable is coronary heart disease indicator, y, and the explanatory variable is age, x. One hundred subjects were included in the study; i.e. $n = 100$. We fit the logistic regression model following the same definition as above, with ω_1 defined for the unknown intercept and ω_2 for the effect of age on heart disease status. Using initial condition $(0, \xi, -1)$ and Barndorff-Nielson format approximation, a two-sided testing p-values is 5.532326×10^{-8}, and five and ninety-five posterior percentiles are of 0.07 and 0.15 respectively.

6. Conclusion

Matching priors were first proposed by Welch & Peers (1963) and Peers (1965). In the general parametrization, if the parameter of interest and the nuisance parameters are not orthogonal, solving the prior from a first order partial differential equation is nontrivial. This paper presents a practical way to solve for the matching priors and the procedure can be suitable to all kinds of initial conditions. Matching priors can be used with the approximations of DiCiccio & Martin (1993). By choosing differential initial conditions one is able to improve the performances of DiCiccio and Martin's approximations.

References

Barndorff-Nielsen, O. E. (1980). Conditionality resolutions. *Biometrika* **67**, 293–310.

Cox, D. R. & Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *J. R. Statist. Soc. Ser. B* **49**, 1–39.

Curran, J. (2005). *Bolstad: Bolstad functions*. R package version 0.2-12.

DiCiccio, T. J. & Martin, M. A. (1993). Simple modifications for signed roots of likelihood ratio statistics. *J. R. Statist. Soc. Ser. B* **55**, 305–316.

Hindmarsh, A. C. (1983). *Scientific Computing*. Amsterdam: North-Holland. Ed. Stepleman, R.W. et al.; Vol. 1 of IMACS Transactions on Scientific Computation.

Hosmer, D. W. & Lemeshow, S. (2000). *Applied Logistic Regression*. New York: Wiley, 2nd ed.

Levine, R. A. & Casella, G. (2003). Implementing matching priors for frequentist inference. *Biometrika* **90**, 127–137.

Lugannani, R. & Rice, S. (1980). Saddlepoint approximation for the distribution of the sum of independent random variables. *Adv. Appl. Prob.* **12**, 475–490.

Peers, H. W. (1965). On confidnece points and bayesian probability points in the case of several parameters. *J. R. Statist. Soc. Ser. B* **27**, 9–16.

Petzold, L. R. (1983). Automatic selection of methods for solving stiff and nonstiff systems of ordinary differential equations. *SIAM: J. Sci. Statist. Comp.* **4**, 136–48.

R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rhee, H., Aris, R. & Amundson, N. (1986). *First-order Partial Differential Equations: Theory and Application of Single Equations*, vol. 1. Englewood Cliffs, New Jersey: Prentice-Hall.

Setzer, R. W. (2007). *odesolve: Solvers for Ordinary Differential Equations*. R package version 0.5-17.

Sweeting, T. J. (2005). On the implementation of local probability priors for interest parameters. *Biometrika* **92**, 47–57.

Tibshirani, R. (1989). Noninformative priors for one parameter of many. *Biometrika* **76**, 604–608.

Welch, B. L. & Peers, H. W. (1963). On formulae for confidence points based on intergrals of weighted likelihoods. *J. R. Statist. Soc. Ser. B* **25**, 318–329.

Zhang, J. (2008). Higher order conditional inference using parallels with approximate bayesian techniques. *Ph.D. Dissertation, Rutgers University* , 40–44.

Table 1: Ratio of two exponential means: type I error probability

	BN Format		LR Format	
Tests	1-sided	2-sided	1-sided	2-sided
Likelihood ratio test	0.0520	0.0526	0.0520	0.0526
I.C. $(0, \xi, -1)$	0.0456	0.0441	0.0456	0.0441
Analytic solution: $1/\psi$	0.0456	0.0441	0.0456	0.0441
I.C. $(0, \xi, -\log \xi)$	0.0499	0.0498	0.0499	0.0498
Analytic solution: $1/(\psi\lambda)$	0.0499	0.0498	0.0499	0.0498

*I.C. stands for initial condition.

†Results are based on 1,000,000 rounds of simulation with $n = 10$.

‡Tests are of nominal type I error 0.05.

Table 2: Logistic regression: type I error probability

	BN Format		LR Format	
Test	1-sided	2-sided	1-sided	2-sided
Likelihood ratio test	0.054	0.060	0.054	0.060
I.C. $(0, \xi, -1)$	0.052	0.057	0.052	0.057
I.C. $[0, \xi, -\log\{(\xi+1)^2+1\}]$	0.028	0.019	0.031	0.020
I.C. $[0, \xi, -\log\{(\xi+1)^{2/5}+1\}]$	0.041	0.041	0.044	0.046
I.C. $[0, \xi, -\log\{(\xi+1)^{2/11}+1\}]$	0.045	0.048	0.046	0.050

*I.C. stands for initial condition.

†Results are based on 10,000 rounds of simulation with $n = 30$.

‡Tests are of nominal type I error 0.05.